Manfred Stede

Shallow - Deep - Robust

Back in 1983, a special issue of the ‚American Journal of Computational Linguistics' was devoted to the processing of „ill-formed input", reflecting the insight of the times that a key to more successful language analysis was *robustness*: the ability of a system to gracefully degrade – rather than crash – when being confronted with unexpected data. Amongst the methods considered were ideas such as relaxing constraints on ATN state transitions in case of parsing failure, or trying to glue together pieces of bottom-up chart parsing results in case no complete analysis was found.

In 1991, the author of these lines wrote a paper that summarized the efforts of the 1980s to achieve robustness. When published as (Stede 1992), it happened to be outdated very soon, because it was so unlucky as to have just missed the „statistical turn" that fundamentally shifted the emphasis of computational linguistics research. And the point of the statistical turn, of course, was to get rid of the robustness problems.

1 The statistical notion of robustness

„Traditional" symbolic language understanding systems were based on rules and constraints that define the precise conditions for an utterance to be well-formed, and therefore to be (in principle, at least) subject to further processing by the system. As is well-known, the inherent problem is that of „scaling up": Hand-crafted models defining the well-formedness conditions have limited coverage, and hence the whole system is bound to fail when input data deviates – ever so slightly – from the model's expectations. The core idea of corpus-based approaches, on the other hand, is to replace hand-crafted well-formedness conditions by a large storage of data, or some intelligent evaluation thereof. Language understanding then is a matter of re-collection: Well-formed, and thus (in principle, at least) subject to further processing by the system are those utterances that have been seen before. Robustness now is at the centre of attention almost by definition, because we can't seriously expect to have seen *every* relevant utterance (*i.e.*, one that the system should be prepared for) before. Re-collection is thus a matter of "best match", with the goal to find the utterance in the data storage that is most similar to the one being processed. This idea is very transparently implemented in example-based machine translation, and even more obviously in speech recognition: A recognizer always comes up with a string of words when being confronted with some auditory input; by definition, it cannot „fail". Underlying the system is a probabilistic model that always selects the most likely output based on what has been seen so far (*i.e.,* what is encoded in the language model). Robustness at its best?

2 Shallow analysis and deep understanding

Besides probabilistic models, a second major approach to creating robust systems in the 1990s was *shallow analysis*: Rather than demanding a complete spanning analysis of each input sentence, one is content with identifying a sequence of smaller constituents, possibly interrupted by un-analysable spans. For certain goals (e.g., 'information extraction'), a shallow analysis can indeed be sufficient, and many interesting applications have been built with this technique. As a general account of language processing, on the other hand, it clearly has its limitations. Determining a sequence of 'chunks' is a means to some end, but not a goal in its own right. In order to actually achieve something with a linguistic utterance in the general case, it needs to be connected to semantics, maybe to knowledge – in some way to the particular world-view of the underlying application. Shallow analysis by definition does not address this, and the aforementioned probabilistic notion of surface-robustness does not help here, either: When the strategy is to produce „rather any analysis than no analysis“, then the analysis might be utterly wrong, as any contemporary speech recognizer demonstrates. A truly robust system would be able to distinguish between „not quite sure, but under the present circumstance this might be the analysis“ and „really no idea“; in the latter case, a 'fail' answer would be better than a useless one that only misleads subsequent processing. The property of robustness, in short, needs to stretch from surface analysis to deeper analysis.

One approach that has received much attention lately is to have both a shallow analysis and a deep processing module and let them run in parallel to compete with each other (*cf.* 'Verbmobil', Wahlster 2000). As deep syntactic parsing has made significant progress with respect to efficiency lately, this competition indeed seems to become feasible. From an engineering perspective, such a model can lead to useful applications with gains in overall efficiency and quality. Nonetheless, it seems that an architecture employing distinct processing engines is rather an interim solution, not chosen for principled reasons but due to a lack of more systematic ways. Constructing a strict dichotomy between shallow and deep processing, along with some ad-hoc mechanism for putting together their results, is a somewhat artificial move that does not offer insights into robust language processing per se.

3 The linguistic beach

When you arrive at the beach and start walking into the water, it is rather unlikely that you will at some specific point feel to cross a line between 'shallow' water and 'deep' water in a single step. With language understanding, I wish to argue now, the picture is just the same. At one end of the scale, when reading or listening, a feeling of complete non-understanding is very rare at best; we usually at least pick up a number of words and their basic connections. Due to our drive to constantly attribute sense to information in context (*cf.* 'Sinnkonstanz', Hörmann 1976), things must get very weird indeed for us to judge something „meaningless“. Then, at the opposite end of the scale, the notion of „full understanding“ is equally questionable. Granted, a semiotic agent such as a traffic light is sufficiently limited in expressivity so we can indeed claim to have it understood *completely* when it turns green. With most linguistic utterances, on the other hand, this is not the case. Even when readers feel comfortable with a text, there will usually be aspects of meaning overlooked, nuances unnoticed, and particular intentions of the author unrecognized. This partiality is not a problem as long as the communication proceeds smoothly. For example, (back to shallow analysis) in order to arrive at the gist of the meaning of an utterance, it often might not be necessary to form a complete and detailed syntactic structure.

For modelling this kind of variable-depth language understanding, the integration of shallow and deep analysis within the same formalism (indeed, the abolition of the dichotomy) seems to be a prerequisite. One example for flexible syntactic analysis along these lines is the „incremental deep parsing“ suggested by (Ait-Mokhtar *et al.* 2002). Of course, flexibility then cannot stop with syntax – semantic analysis also needs to account for under-specification on the sentence level and beyond. This includes mixed-depth representations also for dialogue and for text, as it had been proposed by (Hirst and Ryan 1992), who saw the bare text and its „full“ conceptual representation as the two ends of a spectrum, with many levels in between that represent partiality: unresolved or partially-resolved anaphors, quantifiers and connectives with undetermined scope, lexemes with more than one possible reading, etc. Their point was that for different purposes, portions of utterances may reach a different level on the representation spectrum – and that this is not a fault but a feature.

4 Task-driven, robust analysis

One lesson that was to be learned from the 1990s North American MUC (message understanding conference) competitions was that systems perform best when they are geared to the task at hand, when they look specifically for the information desired rather than attempting broad and general analyses. The lesson of course is not to build a new system for each new domain but to realize that the *task model* should have the power to exercise control over language processing. Specifically, it should be for the task to decide what level of representation is necessary or helpful at some point. For the mixed-depth representations advocated above to be of any use, we also need mixed-depth *processing*: systems that do just as much linguistic work as necessary for moving on with their non-linguistic business. To give an example from machine translation – admittedly not a good example of a 'non-linguistic business' –, (Dorna and Emele 1998) implemented this idea in the „resolution on demand“ strategy of a transfer component that did not perform a particular disambiguation in case the ambiguity could be easily carried over to the target language utterance anyway.

Can this kind of behaviour be generalized systematically? If so, a system would be labelled ‘robust’ if the decision on how much processing takes place were taken away from the completeness constraints operating autonomously at each level of linguistic description, and instead handed over to the underlying *task model:* Once a (partial) representation of an utterance appears to be in line with the overall state of affairs, the contextual conditions, and the goals of the system, then the system can accept it and move on – otherwise, it should do more linguistic work, or in the worst case announce failure. The challenge then is to devise interfaces between task-modules and linguistic modules that allow for flexible propagation of information back and forth (expectations downward, partial interpretations upward). If such mechanisms were then some day standardized and thus portable across applications and domains – computational linguistics would be in very good shape indeed.

5 Summary

Robustness has been a central issue in computational linguistics and – so I argued in this short essay – is very likely to remain one for some years to come, notwithstanding the partial successes that statistical, data-oriented approaches have achieved for surface analysis. Step by step, robust analysis has to be taken from the surface analysis to the deeper levels of processing, and importantly, the property of robustness has to *connect* these levels. When we see language understanding as naturally varying in granularity and depth of analysis, the key issues are representations and processing formalisms for mixed-depth understanding. The factors that determine the depth at some particular point of analysis should be on the one hand the linguistic capabilities of the system, and on the other hand the demands made by the underlying application (or task model). Devising such general mechanisms is not a short-term goal, probably. Still, in the midst of all the (undeniably useful) corpus-oriented and applications-oriented work, it might be worthwhile for computational linguistics to still keep an eye on the „old“ AI-style goals of modelling text understanding and of gaining insights into human language processing by building such models.

-

Literatur

Ait-Mokhtar, S., J.-P. Chanod and C. Roux (2002): Robustness beyond Shallowness: Incremental Deep Parsing. – *Natural Language Engineering* 8 (2/3), 121-144.

Emele, Martin and Michael Dorna (1998): Ambiguity Preserving Machine Translation Using Packed Representations. In: Proc. of the 17th International Conference on Computational Linguistics (COLING-ACL 98), Montréal, Canada, 365-371.

Hirst, Graeme and Mark Ryan (1992): Mixed-Depth Representations for Natural Language Text. In: Paul Jacobs (ed.): Text-Based Intelligent Systems, Hillsdale/NJ: Lawrence Erlbaum Associates.

Hörmann, Hans (1976): Meinen und Verstehen. – Frankfurt: Suhrkamp.

Stede, Manfred (1992): The Search for Robustness in Natural Language Understanding. – *Artificial Intelligence Review* 6, 383-414.

Wahlster, Wolfgang (2000): Verbmobil: Foundations of Speech-to-Speech Translation. – Berlin/Heidelberg/New York: Springer.